Extremes 4:2, 105–127, 2001

Extremal Forex Returns in Extremely Large Data Sets

MICHEL M. DACOROGNA
Zurich Insurance Company, Reinsurance

ULRICH A. MÜLLER
Olsen and Associates, Research Institute for Applied Economics

OLIVIER V. PICTET
Dynamic Asset Management SA

CASPER G. DE VRIES*
Erasmus Universiteit Rotterdam and Tinbergen Institute
Email: cdevries@few.eur.nl

[Received March 24, 2000; Revised and Accepted October 29, 2001]

Abstract. Exciting information for risk and investment analysis is obtained from an exceptionally large and automatically filtered high frequency data set containing all the forex quote prices on Reuters during a ten-year period. It is shown how the high frequency data improve the efficiency of the tail risk cum loss estimates. We demonstrate theoretically and empirically that the heavy tail feature of foreign exchange rate returns implies that position limits for traders calculated under the industry standard normal model are either not prudent enough, or are overly conservative depending on the time horizon.

Key words. extreme value theory, regular variation, large data sets, position limit, foreign exchange rates

AMS Subject Classification. Primary—62G32 62G20

1. Introduction

Asset data sets covering a few thousand price observations per contract are commonly used for financial analysis. Over the past decade high frequency data sets containing tick by tick observations have become available, see Baillie and Dacorogna (1997) who devoted an entire journal issue to the topic of high frequency data in finance. Most of the studies in this area focus on facets related to the center of the distribution and the central limit law. This paper, instead, characterizes the distribution of the outliers at the very highest frequency level in an exceptionally large data set, amounting to over 10 million foreign exchange rate price quotes. Handling such a sizeable data set requires techniques which are novel to economics and finance. The size of the data set enables one to

* Corresponding author. Department of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

accurately determine the probability mass in the tails and demonstrate the uses of extreme value analysis for economic data. We discuss the effects of portfolio diversification under the alternative stochastic models and find that the heavy tails of the typical return distribution induce better diversification properties against the largest risks, than if normality prevailed. Our estimates of the tails are used to calculate reliable overnight position limits for foreign exchange traders over different time horizons. We show theoretically and empirically that extending the time horizon increases the Value-at-Risk (VaR) more rapidly under the normal than under the heavy tail hypothesis, so that the normal model underestimates the short term risk and overestimates longer term risk.[1]

The high frequency data yield a considerable improvement in the efficiency of the position limit estimates. Because outliers by their very nature are rare events, a finely sampled period yields more extreme outcomes (corrected for scale) than if a coarse grid is applied. For a sample of n random variables ε_i, we have that

$$\max\left\{\frac{\varepsilon_1 + \varepsilon_2}{2}, \ldots, \frac{\varepsilon_{n-1} + \varepsilon_n}{2}\right\} \leq \max\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1}, \varepsilon_n\}.$$

This suggests that there could be a benefit from increasing the sampling frequency for information concerning the tailshape. A well known counter-example is the linear estimator of the average return per unit time interval, which does not benefit from an increase in frequency. Because the tail shape estimator is nonlinear in the underlying data one might conjecture that using higher frequency data improves the fit. In the theory section we give conditions and a formal proof to this conjecture; but we also provide counter-examples.[2] The size of the data set is conducive to implementing the Danielsson et al. (2000) subsample bootstrap procedure for selecting the number of tail area observations which must be used in the estimation.

The structure of the paper is further as follows. The next section starts by deriving the implications of heavy tails for portfolio diversification and time aggregation. The statistical benefits from using high frequency data are discussed in Section 3. Section 4 provides a description of the data and data handling. Empirical results on the shape of the tail, the position limits and time aggregation effects are given in Section 5. The last section concludes.

2. Heavy tails and portfolio diversification

We briefly review the statistical properties of the log-returns of foreign exchange rates. The log-return is defined as the natural logarithm of the price ratio p_{t+1}/p_t, where p_t denotes the spot price of one unit of a foreign currency in terms of the domestic currency units. Under continuous discounting, the log-return is the appropriate measure rather than discrete return $p_{t+1}/p_t - 1$; moreover for the log-return Jensen's inequality is not relevant, so that domestic and foreign agents can agree on the moments of the returns. A first property of foreign exchange rate returns (abbreviated as forex returns) is that the mean is approximately zero (when corrected for interest gains[3]). This fair game property is induced by arbitrage activities: If the market gets the smell that tomorrows' exchange rate will be higher, then it pays to buy today and sell back tomorrow, but this very process quickly brings today's price up to the expected price level for tomorrow. Thus the log exchange rate has the martingale property. The second data property is that the forex return empirical distribution function displays more mass in the tails than the normal distribution, which is the standard law used in theoretical finance. The heavy tail property is well documented, see Boothe and Glassman (1987), and Baillie and McMahon (1989). Thirdly, plotting the squares of the forex returns reveals that the returns come in clusters of high and low volatility. The dependence in the second moment of the forex returns is often modeled by means of an ARCH process; see Baillie and McMahon (1989). There is a connection between the volatility clusters and the heavy tails. De Haan et al. (1989) showed that even if the innovations to an ARCH(1) or GARCH(1,1) process are light tailed distributed, say normal, then the unconditional distribution is nevertheless fat tailed distributed in the sense of regular variation see (2) below, and gave solutions for the first order tail index α and the extremal index. The latter quantity is a parameter whose inverse is a measure of the cluster size in extremes stemming from depedency, see Leadbetter et al. (1983). Typically, though, empirical work reveals that one needs to take the innovations to the ARCH process heavy tailed as well in order to obtain a good fit, see Baillie and McMahon (1989, Chapter 4).

Here we elaborate on the fat tail property of the tail observations and show that a number of interesting theoretical economic results can be obtained if one assumes that the return distribution is not normal but heavy tailed. In particular we assume the distribution $F(x)$ satisfies

$$F(x) = 1 - ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})) \quad \text{as}\, x \to \infty, \qquad \alpha, \beta, a > 0. \tag{1}$$

This is somewhat more specific than just assuming that the distribution varies regularly at infinity, i.e.

$$\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \quad \alpha > 0. \tag{2}$$

Regular variation implies the first order Pareto term in (1), via the extreme value theorem. But the second order behavior assumed in (1) is exploited to obtain a number of novel economic insights. Note that (1) applies for example to the Student-t distribution, which is often used as a model for the unconditional distribution of forex returns in empirical modeling, see Boothe and Glassman (1987). In this case the tail index α equals the degrees of freedom. Generally, the tail index corresponds to the number of bounded moments. The parameter a is a scale parameter. The β and b are the corresponding second order parameters (in case of the Student distribution $\beta = 2$).

The class of fat tailed distributions is closed under certain functional operations. Let the X_i be i.i.d. with common distribution $F(x)$ which satisfies (1) and thus (2). For the sum $\Sigma_i^n X_i$ we have by Feller's celebrated theorem (1971, VIII. 8) that asymptotically

$$P\{\Sigma_{i=1}^{n} X_i > x\} = nax^{-\alpha}(1 + o(1)), \quad \text{as } x \to \infty, \tag{3}$$

and where the factor a is as in (1). Thus the tail probabilities are additive for large x. A first implication of this result is for the distribution of the average. For large n, the distribution of $\Sigma_i^n X_i/n$ becomes similar to a normal distribution in the center (by the central limit theorem when $\alpha > 2$). But this finite average still has bounded moments only up to α. Therefore extreme events nevertheless carry a probability quite different from the normal probabilities. This has implications for the analysis of risk and portfolio management.

We first investigate the single period effect on the Value-at-Risk (VaR) if the returns are fat tailed distributed in comparison with the benchmark finance normal model. The VaR is defined as the loss quantile at which

$$P\{X \leq -\text{ VaR }\} = p,$$

where p is the acceptable risk of loss, see Dowd (1998) for the economics behind the VaR concept. Let Y be a standard normal random variate, and let X have a fat tailed distribution which satisfies (2), and the equivalent on the left tail. The following result exploits the power rate from (2) versus the exponential rate of the normal distribution.

Proposition 1: *For any fat tailed distributed X and normal variate Y there is a T such that for all $t > T > 0$*

$$P\{X \leq -t\} > P\{Y \leq -t\}.$$

Sufficiently far out in the tails the presumption of normality understates the risk of loss if the returns are in fact fat tailed distributed. Asymptotically, the ratio $P\{X \leq -t\}/P\{Y \leq -t\}$ even becomes unbounded. While mathematically simple, the result serves as an important reminder to the standard practice of the banking sector to base VaR calculations on the normal distribution.

The next result investigates how an increase of the time horizon over which position limits can be held, affects the VaR for a given risk level p. The following VaR's x and y, where $x, y > 0$, are associated with this risk level

$$P\{-X > x\} = P\{-Y > y\} = p. \tag{4}$$

For the purpose of presentation we have mapped the left tail into the positive orthant. Suppose we want to know how time aggregation, i.e. $\sum_i^n X_i$ and $\sum_i^n Y_i$, affects the VaR. Commercial banks often calculate a one day VaR level for internal risk management purposes, but also have to report a 10-day VaR level for regulatory purposes. The ten day period is chosen to take care of difficulties related to unwinding large and illiquid positions. Let q and t denote the respective proportional changes in the VaR that keep the risk level p unchanged:

$$P\left\{-\sum_{i=1}^{n} X_i > qx\right\} = P\left\{-\sum_{i=1}^{n} Y_i > ty\right\} = p. \tag{5}$$

We have the following somewhat surprising result.

Proposition 2 (The α-root rule): *Suppose X has finite variance, so that* $\alpha > 2$. *At a constant risk level p, increasing the time horizon increases the VaR for the normal model percentage-wise by more than for the fat tailed distribution, i.e.* $t > q$, *for all p sufficiently small.*

Proof: Under the normal model in order to guarantee that

$$p = P\{-Y > y\} = P\left\{-\sum_{i=1}^{n} Y_i > ty\right\},$$

we take $t = \sqrt{n}$. For fat tailed distributions for small p, i.e. with x sufficiently large, we have by (1) and (3) that for $q = n^{1/\alpha}$

$$p \approx P\left\{-\sum_{i}^{n} X_i > n^{1/\alpha}x\right\} = ax^{-\alpha}(1 + o(1)) \text{ as } x \to \infty.$$

If X has finite variance, then $\alpha > 2$. Hence, with $1/2 > 1/\alpha$ we see that $t > q$. □

In practice risk managers often employ the square root of time rule to translate single day VaR calculations into the regulatory required 10-day VaR, to circumvent the extra work of explicitly estimating the 10-day VaR. The above proposition shows that, if the unconditional return distribution has at least a bounded second moment, increasing the time horizon increases the VaR for the fat tailed distributed returns by a smaller percentage than if the returns were normal distributed. This means on the one hand that the presumption of normality may yield too conservative capital adequacy requirements for the multi-day risk horizon. Stated otherwise, larger speculative positions would be permissible. On the other hand, as the first proposition shows, the normal based one-day VaR calculations can be too lax.[4] A caveat to the result is the fact that in the real data there is some second moment dependence, such that the α-scaling may not be entirely correct either.

The above propositions can also be used to investigate the benefits from cross-sectional portfolio diversification. Suppose the following standard one factor capital asset pricing model (CAPM) applies

$$R_i = \eta_i R + Q_i.$$

Here R is the excess return on the market portfolio (in excess of the risk free rate), Q_i is the zero mean idiosyncratic risk. The idiosyncratic risk can be reduced through diversification. The η_i is the ratio of the covariance between R_i and R, and the variance of R. It indicates the risk premium $\eta_i E[R]$ commanded by asset i for investors to be willing to hold its supply at the going market price. Consider an equally weighted portfolio of m assets (weights $1/m$).

Corollary 3 (Diversification benefits): *Suppose the Q_i are i.i.d. distributed with $P\{-Q_i > x\} = ax^{-\alpha}(1 + o(1))$ as $x \to \infty$, and that $P\{-R > x\} = o(P\{-Q_i > x\})$. For large loss levels the tail diversification benefits from the equally weighted portfolio are larger if the returns have finite variance but are fat tailed, than if they are normally distributed.*

Proof: For large x by arguments from the proof to Proposition (2)

$$P\left\{\frac{1}{m}\sum_{i}^{m} R_i \leq -x\right\} = P\left\{\frac{-1}{m^{1-1/\alpha}} Q_i > x\right\}(1 + o(1)) \tag{6}$$

$$= am^{\alpha+1}x^{-\alpha}(1 + o(1)). \tag{7}$$

Since $\alpha > 2$ for the fat tail distributed returns, in contrast to the case of normal distributed returns when $\alpha = 2$ on the LHS of (6), the claim follows. □

Diversification is more effective if returns are fat tailed distributed than if the underlying distribution is normal. It has been noted in the economics literature that the effect of diversification is less pronounced if $\alpha < 2$ in comparison with the normal distribution. Fama and Miller (1972, p. 270) discuss the case of sum stable distributions. They note that for $\alpha < 1$ diversification actually increases the dispersion. We are not aware of a discussion of the case $\alpha > 2$. The reason for this might be that when $\alpha > 2$ the variance is finite and hence the central limit law kicks in. But this normality result only applies to the center of the distribution. For the extreme loss levels diversification is more effective if the return distributions are fat tailed with $\alpha > 2$, than if the returns are normally distributed. *Ipso facto* the upside potential is also reduced by a larger factor.

3. Estimation and high frequency benefits

The empirical economic issue of the paper is the determination of the overnight position limits for a forex dealer. Position limits indicate how large the (speculative) open positions in foreign currencies can be, see Dowd (1998) for a textbook treatment of current industry practices. Often a worst case scenario analysis is used to determine the position limits. Note that the worst case is just the VaR at $p = 1/n$. It is also one of the methods that can be used to compute the regulatory requirements imposed by the bank for international settlements (BIS) on the size of open positions. The disadvantage of this common practice is that one is unable to cope with risk levels below the inverse of the sample size $1/n$, and that the fully non-parametric method is inefficient and biased. As many in the banking industry have come to realize, the semi-parametric approach which exploits the data feature of regular variation (1) allows one to achieve greater precision, enables extrapolation beyond the sample minimum, while it is not burdened by the straightjacket of a fully parametric model.

To determine the position limit, we need an estimate of the VaR at the risk level p. We use the simple quantile estimator from De Haan et al. (1994). For $p < 1/n$ we estimate x_p by

$$\hat{x}_p = X_{(m)} \left(\frac{m}{np} \right)^{\hat{\gamma}}, \tag{8}$$

where $\hat{\gamma}$ is an estimate of $1/\alpha$, the $X_{(i)}$ are the descending order statistics from the sample $\{-X_1, \ldots, -X_n\}$, $1 \leq m < n$, and where for ease of presentation the losses are mapped into the positive half axis. Hence, if the bank specifies a probability level p, where p is borderline in-sample or out of sample, then the above method allows us to compute the position limit $\hat{x}_p$. In (8) one needs an estimate of the tail index. To estimate $\hat{\gamma}$ we use the Hill (1975) estimator:

$$\widehat{\gamma} = \frac{1}{m} \sum_{i=1}^{m} \ln X_{(i)} - \ln X_{(m+1)}. \tag{9}$$

The tail index estimator (9) and the quantile estimator (8) require a choice of the nuisance parameter m, $1 \leq m < n$, which determines how many of the highest order statistics are taken into account. Traditionally the ''start of the tail'' has been determined by graphical inspection of the Hill and quantile plots, whereby $\widehat{\gamma}$ and $\hat{x}_p$ are plotted against m to locate the point where the variance and bias contribution to the mean squared error are about equally small. The asymptotic mean squared error (AMSE) of $\widehat{\gamma}$ when (1) applies reads

$$\text{AMSE } (\hat{\gamma}, m) = \frac{1}{\alpha^2} \left(\frac{1}{a} \right)^{\frac{2\beta}{\alpha}} \frac{\beta^2 b^2}{(\alpha + \beta)^2} \left(\frac{m}{n} \right)^{\frac{2\beta}{\alpha}} + \frac{1}{\alpha^2} \frac{1}{m}, \tag{10}$$

where the first part is the biased squared, and the second part is the variance, see Goldie and Smith (1987). Minimize this AMSE with respect to m to find the optimal number $\overline{m}$ of extreme order statistics

$$\overline{m} = un^{2\beta/(2\beta+\alpha)}, \quad u = \left(\frac{\alpha(\alpha+\beta)^2}{2\beta^3 b^2} a^{2\beta/\alpha} \right)^{\alpha/(2\beta+\alpha)},$$

and where the AMSE is

$$\text{AMSE } (\hat{\gamma}, \overline{m}) = qn^{-2\beta/(2\beta+\alpha)},$$

$$q = \frac{2\beta + \alpha}{\alpha^3}\left(\frac{\beta}{\alpha+\beta}\right)^{2\alpha/(2\beta+\alpha)}\left(\frac{\alpha}{2\beta a}\right)^{2\beta/(2\beta+\alpha)}\left(b^2\right)^{\alpha/(2\beta+\alpha)}.$$

Note that for $m = \overline{m}$ the bias part and the variance part of the AMSE in (10) vanish at the same rate: $n^{-2\beta/(2\beta+\alpha)}$. If either the bias or the variance part is converging at a higher rate, then the other part converges more slowly than $n^{-2\beta/(2\beta+\alpha)}$, and hence the whole AMSE converges more slowly. It follows that it is optimal to have the bias squared converging to zero at the rate of the variance, see Hall (1982) and Goldie and Smith (1987). While graphical methods can be used to locate $\overline{m}$ visually, this is a somewhat arbitrary process. Recently Danielsson et al. (2001) and Danielsson and De Vries (1998), have developed a subsample bootstrap procedure which delivers a consistent estimate of $\overline{m}$; see Drees and Kaufman (1998) for an alternative procedure. Subsampling is needed to achieve convergence in probability. To be able to implement such a procedure one needs to construct subsamples which are, on the one hand, an order of magnitude smaller than the full sample size. On the other hand, because the outliers are rare by their very nature, one needs subsamples which are still quite sizable. The very large data set of Olsen and Associates is therefore very instrumental for the subsample bootstrap procedure.

Before turning to the application, we return to the claim made in the introduction that the high frequency data generally improve the MSE properties of $\hat{\gamma}$. We consider the case $\alpha > 2$ only, since this is the empirically relevant configuration. In this case both the mean and the variance are finite, but the moments larger than α are unbounded. We first obtain a general lemma on second order convolution behavior. This result is needed because the AMSE is a function of the first and second order parameters, cf. (1) and (10). The published literature only gives a result on second order convolution behavior for positive random variables, see Geluk et al. (1997). But since the log-asset returns can be positive and negative, we need to analyze this case afresh. To restrict the number of different combinations that can arise, we assume that the tails are symmetric. Symmetry is a natural property of the forex return data if countries pursue similar monetary policies, see Muller et al. (1990). We find that because the distribution of asset returns is two-sided, and as in the case of forex returns $E[x] = 0$, a new factor related to $E[x^2]$ enters.

Lemma 4 (Second order convolution): *Suppose that the tails are second order similar, i.e. as $x \to \infty$*

$$P\{X \le -x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})), \tag{11}$$

$$P\{X > x\} = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})),$$

and where $a > 0$, $b \neq 0$ and $\beta > 0$. Moreover, assume that $\alpha > 2$, so that $E[X]$ and $E[X^2]$ are finite. Suppose X_1 and X_2 are i.i.d. and satisfy (11).

Then for the 2-convolution as $s \to \infty$

$$
\begin{aligned}
&P\{X_1 + X_2 > s\} \\
&\quad = 2as^{-\alpha}\left(1 + bs^{-\beta} + \alpha E[X]s^{-1} + \frac{\alpha(\alpha+1)}{2}E[X^2]s^{-2}\right) \\
&\qquad + o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}).
\end{aligned} \tag{12}
$$

Proof: See Appendix A. □

Remark 1: *Note that by symmetry*

$$
\begin{aligned}
&P\{X_1 + X_2 \leq -s\} \\
&\quad = 2as^{-\alpha}\left(1 + bs^{-\beta} - \alpha E[X]s^{-1} + \frac{\alpha(\alpha+1)}{2}E[X^2]s^{-2}\right) + o(s^{-\alpha-2}) \\
&\qquad + o(s^{-\alpha-\beta})
\end{aligned}
$$

where in view of (12) there is now a negative sign on the $\alpha E[x]s^{-1}$ term.[5] *To see this directly, use $P\{X_1 + X_2 \leq -s\} = P\{-(X_1 + X_2) > s\}$, while by the tail symmetry $P\{-X > x\} = P\{X > x\}$ but $E[-X] = -E[X]$.*

Proposition 5: *Suppose the X_i are i.i.d. with a distribution $F(x)$ that is symmetric around zero and is regularly varying with $\alpha > 2$. Then a w-convolution affects the AMSE$(\hat{\gamma})$ as follows*:

1. $\beta < 2$. *There is no effect*;
2. $\beta = 2$. *The AMSE$(\hat{\gamma})$ changes by a factor*

$$
\left[\left(1 + \frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]/b\right)^2\right]^{\alpha/(4+\alpha)};
$$

3. $\beta > 2$. *The AMSE$(\hat{\gamma})$ changes by a factor*

$$
c\left[\frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]\right]^{\frac{2\alpha}{4+\alpha}}\left(\frac{1}{b^2}\right)^{\alpha/(2\beta+\alpha)},
$$

and where

$$
c = \frac{4+\alpha}{2\beta+\alpha}\left(\frac{2}{\alpha+2}\right)^{\frac{2\alpha}{4+\alpha}}\left(\frac{\alpha+\beta}{\beta}\right)^{\frac{2\alpha}{2\beta+\alpha}}\left(\frac{\alpha}{4an}\right)^{\frac{4}{4+\alpha}}\left(\frac{2\beta an}{\alpha}\right)^{\frac{2\beta}{2\beta+\alpha}}.
$$

Proof: The symmetry around zero assumption coupled with $\alpha > 2$ imply that $E[X] = 0$ and $E[X^2]$ finite. Let s be sufficiently large and iterate on (12). This gives:

1. $\beta < 2$. $P\{\sum_i^w X_i \leq s\} \approx 1 - was^{-\alpha}[1 + bs^{-\beta}]$.
2. $\beta = 2$. $P\{\sum_i^w X_i \leq s\} \approx 1 - was^{-\alpha}[1 + \{b + \frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]\}s^{-2}]$.
3. $\beta > 2$. $P\{\sum_i^w X_i \leq s\} \approx 1 - was^{-\alpha}[1 + \frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]s^{-2}]$.

Subsequently, the AMSE($\hat{\gamma}$)s can be compared by using the expression from (10) and by taking ratios:

1. $\beta < 2$. In the expression for the AMSE the sample size n changes by the frequency factor $\frac{1}{w}$, but this cancels against the effect on a. Because b and β are unaffected, there is no change.
2. $\beta = 2$. Again, the effects on n and a cancel. But b is affected, and this yields the factor in the claim.
3. $\beta > 2$. In this case, not only do a and b change, but the second order tail index β also changes as it collapses to 2. □

The upshot of Proposition 5 is that either time aggregation has no effect, i.e. when $\beta < 2$, or that the AMSE($\hat{\gamma}$) deteriorates, possibly only after the first few convolutions. If $\beta > 2$ the AMSE($\hat{\gamma}$) always deteriorates after the first convolution. If the second order scaling constant $b < 0$ and $\beta = 2$, then it is possible that the AMSE($\hat{\gamma}$) improves at the first few convolutions. While it can thus not be ruled out that higher frequencies deteriorate the AMSE properties of $\hat{\gamma}$ for the first few convolutions, the majority of the cases goes into the other direction. It is therefore of interest to see what the data have to say on this matter.

4. Data set construction and data filters

The foreign exchange market is a worldwide market with no business hours limitations. The bid and ask offers of major financial institutions (market makers) are conveyed to customers' screens by large data suppliers and the deals are negotiated over the telephone or electronically. The quoted prices are not actual trading prices, although they are binding for serious financial institutions (given some volume restrictions). Our data set has been described in Müller et al. (1990) and Dacorogna et al. (1993), so the description here is limited to the relevant points for the study of extreme values.

The first point is that bid quotes are more reliable than the ask quotes, since the data contain the bid price as a full number, but only give the last two digits of the corresponding ask price. Therefore we study the logarithmic bid price x_j^{bid}, which is defined as $x_j^{\text{bid}} \equiv \ln p_{\text{bid},j}$.

The second point is that data filtering is necessary, to correct for technical and human errors in the data processing stages. The data supplier transmits the data in the form of text. A typical error is wrong decimal digits due to failed text updates. This is identified by the so-called decimal error filter. The resulting raw time series of price records is subsequently filtered by a real-time filter which rejects prices that are very unlikely to be serious quotes. The real-time filter algorithm is composed of two parts:

1. The bid price filter which considers a quote with index j to be valid if the following two conditions are fulfilled:

$$\begin{aligned} &\text{i.} \quad \left|\Delta \hat{x}^{\text{bid}}_{j,j'}\right| < Q, \\ &\text{ii.} \quad \left|\Delta \hat{x}^{\text{bid}}_{j,j'}\right| < S \quad s_{j'} + A \Delta t^{D}_{j,j'}, \end{aligned} \tag{13}$$

   where $\Delta \hat{x}^{\text{bid}}_{j,j'} = \hat{x}^{\text{bid}}_j - x'^{\text{bid}}_j$, $\hat{x}^{\text{bid}}_j$ is the logarithm of the jth bid price being validated, and $x^{\text{bid}}_{j'}$ the logarithm of the last valid bid price before $\hat{x}^{\text{bid}}_j$. The $s_{j'}$ is the logarithmic spread of the last valid price as defined in Müller et al. (1990), $\Delta t_{j,j'}$ is the time between the prices $\hat{x}^{\text{bid}}_j$ and $x^{\text{bid}}_{j'}$ expressed in units of days, and Q, S, A and D are the filter parameters, see Table 1.
2. The spread filter which considers a quote to be valid if the bid-ask spread satisfies the following conditions:

$$\begin{aligned} &\text{i.} \quad V < \hat{s}_j < W, \\ &\text{ii.} \quad \left|\ln \frac{\hat{s}_j}{s_{j'}}\right| < C + T \Delta t^{D}_{j,j'}, \\ &\text{iii.} \quad \left|\ln \frac{\hat{s}_j}{s_{j'}}\right| < U, \end{aligned} \tag{14}$$

   where $\hat{s}_j = \hat{x}^{\text{ask}}_j - \hat{x}^{\text{bid}}_j$ is the logarithmic spread of the tested price as defined in Müller et al. (1990). A price must be accepted by both the bid and the spread filter to be valid. The filtering parameters reject never more than 0.6% of the data.

The tail statistics study requires data in the form of a time series which is equally spaced in time. As in Müller et al. (1990) and Dacorogna et al. (1993), linear interpolation over time is used to determine price values within data holes and to generate regularly spaced time series. An hole is said to occur if the distance from one of the two interval limits to the nearest tick is larger than the interval size. The hole rate can be rather high. For example, for 10-minute returns the dollar-Deutsche mark has a 30% hole rate. But this fact does hardly affect our analysis. The holes occur mostly over weekends and in other inactive

Table 1. Filter parameters.

	S	A	D	Q	C	T	U	V	W
FX rates	2.2	0.27	0.6	0.4	1.5	75.0	5.5	$8 * 10^{-5}$	0.04

Table 2. Data frequency.

Values	30 minutes	1 hour	2 hours	6 hours	1 day
w	1	2	4	12	48
N	175,343	87,672	43,836	14,611	3652
n	157,806	78,903	39,447	13,149	3285
n_1	7917	4707	2799	1227	433
n_2	397	280	198	114	57
n_{ov}	1	1	1	3	6

Notes. The table reports the time aggregation factor or convolution factor w, the total number of data N, the sample size n and two subsample sizes. If overlapping intervals are used, as is the case for the 6 hour and 1 day returns, the number of observations is higher; the number of overlapping time intervals is reported in the last row.

market periods and occur rarely during the most active trading hours when extreme price movements usually occur. The tail estimation algorithm, however, only uses these extreme moves and hence is not much affected by the holes.

The sampling period contains ten full years of intra-day data from January 1 1987, 00:00:00 (midnight GMT) to December 31 1996, 24:00:00 (midnight GMT). We validated the stationarity of the sample through reestimation by deleting parts of the sample. The tail shape appears to be quite invariant to time deletion. To obtain aggregated data, returns over larger time intervals are generated from the highest frequency data. We choose these larger intervals as multiples of the basic time interval of 10 minutes (the 10 minute interval is the minimum to filter out irrelevant noise due to e.g. the speed by which different traders update their quotes). In Table 2, the number of observations are given depending on the time interval of the series. To cope with the data reduction at the 6 hour and 1 day horizon, we also created overlapping return data by using a moving window of respectively 2 hours and 4 hours, yielding returns with respectively a two-fold and five-fold overlap; this is indicated in the last row of Table 2. The point estimates hardly differ depending on whether overlapping data or non-overlapping returns are used, but the results based on the non-overlapping return data were of course more variable. The tables which follow are all based on the overlapping return series.

5. Empirical analysis

This section first provides simulation results for some theoretical distributions concerning the tail shape estimator, and subsequently reports the empirical values for a number of forex contracts. Estimates for the position limits are discussed in the second part of the section.

5.1. *Empirical tail shape*

We simulated data from the non-normal symmetric stable distributions with characteristic exponents equal to 1.25, 1.50 and 1.75; the Student-*t* distributions with degrees of freedom equal to 3.00, 4.00 and 6.00; and an ARCH(1) process. The ARCH(1) process has conditional standard normal innovations; in the volatility equation the parameter for the intercept is $1{*}10^{-9}$, and the AR coefficient is 0.97. The tail index α for these models is respectively equal to the characteristic exponent, the degrees of freedom and approximately 2.0848. Where the last tail index for the ARCH process is obtained as the numerical solution to $\Gamma(\alpha/2 + 1/2) = \sqrt{\pi}\big(2 * 0.97\big)^{-\alpha/2}$, see De Haan et al. (1989).[6] The second order tail index β equals α for the stable distributions, while $\beta = 2$ for the Student class. For the ARCH model, from the work of Goldie (1991) we have the lower bound 1 for β (note that Goldie gives the O-bound rate for the square of the process; also note that if it could be shown that the second order expansion (1) applies, then $\beta = 2$).

The estimates are obtained with the subsample bootstrap method, where we used $\varepsilon = 0.25$ for the subsample bootstrap size $n_1 = n^{1-\varepsilon}$.[7] Given the wealth of data, we decided to use a jackknife procedure for calculating the point estimates. To this end the

Table 3. Simulations for the tail exponents.

df	True	Parameters	30 minutes	1 hour	2 hours	6 hours	1 day
Stable	1.25	α	1.28	1.29	1.29	1.23	1.18
			±0.02	±0.03	±0.04	±0.06	±0.10
	1.00	β/α	1.74	1.65	1.61	1.52	1.33
Stable	1.50	α	1.73	1.72	1.70	1.64	1.54
			±0.03	±0.04	±0.05	±0.08	±0.13
	1.00	β/α	1.76	1.67	1.57	1.46	1.25
Stable	1.75	α	2.54	2.53	2.57	2.52	2.50
			±0.05	±0.07	±0.09	±0.15	±0.32
	1.00	β/α	1.54	1.42	1.33	1.14	0.87
Student	3.00	α	2.91	3.12	3.41	3.85	4.79
			±0.20	±0.23	±0.38	±0.45	±1.69
	0.66	β/α	0.70	0.73	0.66	0.66	0.44
Student	4.00	α	3.80	4.35	4.73	5.83	7.44
			±0.33	±0.57	±0.82	±1.44	±5.54
	0.50	β/α	0.61	0.55	0.50	0.46	0.31
Student	6.00	α	5.21	6.14	7.65	8.38	14.31
			±0.52	±1.03	±1.64	±2.49	±18.18
	0.33	β/α	0.55	0.43	0.43	0.39	0.21
ARCH(1)	2.08	α	1.98	1.98	1.93	2.07	2.27
			±0.07	±0.13	±0.11	±0.22	±0.37
	*	β/α	1.03	0.86	1.02	0.85	0.80

Notes. The table gives tail index estimates for simulated data. *The theoretical lower bound for β/α is $1/2.08 \approx 0.48$. Details behind the table are provided in the main text.

data were sliced into 10 equal parts, and estimation proceeded by omitting one slice at a time and subsequently taking the average of the ten different estimates. In total this was repeated only ten times for each distribution, since by then the jackknife and theoretical standard errors $1/\alpha\sqrt{m}$ were already very similar. Furthermore, since the distributions are symmetric, we report averages of both tail estimates (averaged over the ten simulations). Table 3 gives the results. The first and third rows give the true values and the estimated α and β/α values. The second row gives the 2-standard error confidence band based on $1/\alpha\sqrt{m}$. Since we only have consistency of the β/α estimator, no standard errors are reported for this ratio.

Concentrating first on the highest frequency results in the 30 minutes column, we note that the first order tail index estimates are reasonably close, even for the dependent data from the ARCH process. The standard error band is rather small. The ratio of the first to the second order tail index β/α is less precise, but still reasonable. For the ARCH process the β/α ratio is unknown, but appears to hover around 1. So the rate by which the second order term in (1) vanishes appears twice as high as the lower bound given by Goldie (1991).

There is bias in the first order tail index estimates as the stable and Student distributions approach the normal (which lies at the end of the spectrum of both classes of distributions; i.e. as the characteristic exponent approaches 2, respectively as the degrees of freedom becomes unbounded[8]). The sign of b explains the biases in the Student and stable laws when these are close to the normal. The amount of bias is traded of against the amount of variance by the MSE criterion.

Going to the time aggregated data we do find a confirmation for Lemma 4. The downward bias in the one period (30 minutes) α estimates for the Student distribution, turns into an upward bias on the convoluted data. This is in perfect agreement with the theory, since for the Student-t distribution $b < 0$, while the second order scale parameter b is always positive due to time aggregation in case of Student random variables (use (12)). The bias of $\widehat{\gamma}$ has sign $(\widehat{\gamma}) =$ sign $(-b)$, thus the bias switches sign as well. Per contrast the sign of b for the stable laws remains positive and hence the upward bias persists under time aggregation.

Another effect that is readily apparent from the table, is that the bias and standard errors for $\widehat{\alpha}$ increase after the first or second convolution. On the basis of Proposition 5 we obtain predictions for the factor by which the AMSE deteriorates as we time aggregate the data. For the Student-t with 3 degrees of freedom these factors are 10, 31, and 107 respectively for the 2, 6 and 24 hour time horizons, using the 30 minute (unconvoluted) data as the benchmark. From the experimental data in Table 3 we calculate these factors as 11, 42 and 218 respectively. Apparently, at the 1d level, where we use only 3652 observations, see Table 2, there is quite a discrepancy between the theoretical and simulated MSE, but with more data the ratio of the theoretical and simulated MSE comes close to one. Hence the importance of high frequency data.

In Table 4 we report tail index estimates for several dollar exchange rates. The estimates of α all hover around 3.3. From Table 4 we see that there is a tendency for α to increase as the return horizon is increased. This corroborates again Lemma 4, and the convolution formulas under (2) and (3) in the proof to Proposition 5. For example, the upward bias in $\widehat{\alpha}$ increases if the returns are time aggregated and if b in (1) is positive and $\beta = 2$. But the

Table 4. Tail index estimates for different dollar rates.

Rate	Parameters	30 minutes	1 hour	2 hours	6 hours	1 day
USD-DEM	α	3.27	3.32	3.46	4.59	4.74
		± 0.36	± 0.34	± 0.54	± 1.42	± 2.15
	β/α	0.55	0.57	0.52	0.42	0.40
USD-JPY	α	3.44	3.50	3.56	4.20	3.98
		± 0.39	± 0.42	± 0.59	± 1.30	± 1.71
	β/α	0.54	0.55	0.50	0.41	0.40
GBP-USD	α	3.48	3.77	3.63	5.34	5.06
		± 0.58	± 1.20	± 0.82	± 2.14	± 2.46
	β/α	0.46	0.39	0.47	0.33	0.34
USD-CHF	α	3.46	3.75	3.80	5.12	4.59
		± 0.36	± 0.60	± 0.83	± 1.98	± 2.23
	β/α	0.54	0.47	0.43	0.36	0.39

Notes. The table gives the first and second order tail index estimates and the 2-standard error band for the first order tail index estimate; estimation method and parameters are as for Table 3. Symbols for the various contracts are as used by Reuters.

increases in the first order tail index estimates are sufficiently small to conclude that α is invariant under time aggregation, cf. Lemma 4. For reasons that are similar to the deterioration of the bias, the AMSE will deteriorate under time aggregation if $b > 0$. Therefore we advocate to proceed by estimating the tail indices on the highest frequency data available, and extrapolate the results to the lower frequency data without reestimation of the parameters.

A priori it is plausible that b is positive to begin with, because the 30 minute data are already time aggregated themselves. Taking the product of the α and β/α estimates from Table 4, one finds values for β which hover around 2. There is some reason to believe that for freely floating rates $\beta = 2$. Under freely floating rates and with a similar monetary policy in operation on both sides, the log-return series will be symmetrically distributed. If the symmetry in the density $f(x)$ is due to $f(x) = g(x^2)$, i.e. the argument of $f(x)$ contains the argument x in squared form, such as is the case for the Student density, then generically $\beta = 2$.

5.2. *Position limits*

Our estimates of the tails are used to calculate overnight position limits for forex traders over different time horizons; see Hols and De Vries (1991) for an early empirical analysis of this problem. The position limits of a trader curtail the size of the open position of a trader, putting an effective limit to the trader's VaR. The methodology can also be easily adapted to calculate the 10-day regulatory VaR measure. In Tables 5 and 6 we report estimates of the loss quantiles at different risk levels. The risk of loss or probability on a loss is given in the top row of each table as an event that will happen only once per k years (indicated as $1/k$y). The loss estimates are reported for different return horizons. The loss

Table 5. Extreme loss estimates ($\hat{x}_p$) on open positions.

Frequency	1/1 year	1/5 years	1/10 years	1/15 years	1/20 years	1/25 years
USD/DEM						
30 minutes	1.28	2.10	2.60	2.94	3.21	3.44
	(0.33)	(0.37)	(0.39)	(0.40)	(0.41)	(0.42)
1 hour	1.47	2.40	2.97	3.36	3.67	3.94
	(0.44)	(0.48)	(0.50)	(0.51)	(0.52)	(0.53)
2 hours	1.61	2.57	3.15	3.54	3.85	4.11
	(0.60)	(0.66)	(0.69)	(0.71)	(0.73)	(0.74)
6 hours	1.98	2.85	3.36	3.71	3.98	4.20
	(0.97)	(1.08)	(1.13)	(1.17)	(1.19)	(1.21)
1 day	2.37	3.33	3.87	4.23	4.51	4.73
	(1.71)	(1.99)	(2.13)	(2.22)	(2.28)	(2.33)
USD/JPY						
30 minutes	1.15	1.84	2.25	2.53	2.75	2.94
	(0.32)	(0.37)	(0.39)	(0.40)	(0.41)	(0.42)
1 hour	1.29	2.05	2.50	2.81	3.06	3.26
	(0.44)	(0.48)	(0.50)	(0.51)	(0.52)	(0.53)
2 hours	1.50	2.36	2.88	3.23	3.51	3.74
	(0.59)	(0.66)	(0.71)	(0.71)	(0.72)	(0.73)
6 hours	1.79	2.67	3.18	3.53	3.80	4.02
	(0.97)	(1.08)	(1.13)	(1.16)	(1.18)	(1.20)
1 day	2.31	3.47	4.15	4.61	4.97	5.27
	(1.70)	(1.98)	(2.12)	(2.20)	(2.27)	(2.32)

Notes. The table gives extreme losses over different time intervals (30 minutes, 1 hour, 2 hours, 6 hours, 1 day) and corresponding to different loss probabilities (risk). The loss probabilities are expressed as once per a number of *k* years. The values within the parenthesis are the extreme losses computed for a Gaussian distribution with the mean and variance of the data. All values are given in percentages.

estimates are based on (8), in combination with the tail index estimates from Table 4. Within parenthesis we also report the extreme losses computed for a Gaussian distribution with the mean and variance as in the data. All loss estimates are given in percentages.

For the interpretation of the data we use the following example. A typical forex contract has the size of 10 million US dollars. A trader in Frankfurt who holds an open dollar position of that size for one hour runs the risk of loosing at least \$147,000 once per year (Table 5 USD/DEM contract, second row, first column). Once per 25 years the loss on this position will at least be \$394,000. If, instead, the returns are presumed to be normally distributed, the predicted loss levels drop to respectively \$44,000 and \$53,000. Note that under normality the loss levels increase much slower as we go deeper into the left tail than if the return distribution is heavy tailed. The normal model based risk analysis is therefore insufficiently prudent for short horizon losses. From Tables 5 and 6 we also see that the loss levels for the different dollar contracts are very similar.

To give an idea about the size of the position limits implied by the Tables, assume that a bank allocates \$10 million of its own capital to the forex desk. Suppose there are 10 forex traders, i.e. there is \$1 million available per trader. Suppose the bank is willing to risk losing \$1 million per trader. For the USD/DEM contract and a risk tolerance of once per 10 years, the position limit for a half hour trading horizon amounts to \$38.5 million per trader (\$1 million/2.60%). At the daily horizon the position limit is reduced to \$25.8 million. For other contracts we find very similar results.

The tables also show the scope for the α-root rule from Proposition 2 and the relevance of the Lemma 4 on second order precision. In the Tables 5 and 6 we report loss estimates at different horizons for probabilities expressed as events which happen only once per k years. A moments reflection shows that for a given column $1/k$y, the probabilities of the events do increase by the same factor as by which the time horizon is increased (frequency is decreased). A daily event has a 48 times higher probability of occurring once per year than a 30 minute event that occurs only once per year. On the basis of the α-root rule, recall Feller's formula (3), one would therefore expect the quantiles to remain constant if we move down in a particular column, and we could have refrained from reporting at different horizons. Given that the precision of the tail index estimates is likely to be highest at the highest frequency data, cf. Proposition 5, only reporting the highest frequency estimates from the first rows would then seem to be the best strategy. But as the other rows in the Tables 5 and 6 show, the column-wise estimates are not constant.

Just reporting the first row estimates would be an acceptable method if the second order tail parameter $\beta < 2$, or if the increase in the time horizon is small. If $\beta \geq 2$, then Lemma 4 and the expressions under 2 and 3 in the proof to Proposition 5 show that the second order term is affected by the time aggregation. Moreover, if the data are not independent, then time aggregation affects the scale a by a factor which does not necessarily equal the number of convolutions. By estimating the quantiles at the time aggregated data, as we did in the construction of the Tables 5 and 6, the second order effect is implicitly taken into account. Since the dollar contracts yield estimates of β which hover around 2, time aggregation increases the second order term. If $\beta = 2$ the high frequency (30 minutes) loss s_1, say, approximately carries a probability $as_1^{-\alpha}[1 + bs_1^{-2}]$, corresponding to an occurrence risk of $1/k$ years. The w times convoluted loss event s_w, say, that occurs once per k years, carries a probability of $was_w^{-\alpha}[1 + \{b + \frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]\}s_w^{-2}]$; see 2 in the proof to Proposition 5. Since the low frequency event has a probability which is w times as large as the high frequency event (within the given time span of k years), this low frequency probability must also be equal to $was_1^{-\alpha}[1 + bs_1^{-2}]$.[9] It follows that

$$s_1^{-\alpha}[1 + bs_1^{-2}] = s_w^{-\alpha}\left[1 + \left\{b + \frac{1}{2}\alpha(\alpha+1)(w-1)E[X^2]\right\}s_w^{-2}\right]. \tag{15}$$

Hence for the loss quantiles we must have that $s_w > s_1$, in order to be able to equate the two expressions (if b is positive or not too negative). This increase in quantiles is what we see in the Tables 6 and 7 if we move downwards in a particular column.

The equality (15) can also be used to construct long horizon predictions for which one has insufficient data available for reliable direct estimation. For example, take the USD/DEM contract and plug into (15) the 30 minute and 2 hour quantile estimates from Table 5

Table 6. Extreme loss estimates ($\hat{x}_p$) on open positions.

Frequency	1/1 year	1/5 years	1/10 years	1/15 years	1/20 years	1/25 years
USD/GBP						
30 minutes	1.15	1.83	2.24	2.51	2.73	2.91
	(0.31)	(0.35)	(0.37)	(0.38)	(0.39)	(0.40)
1 hour	1.30	2.00	2.40	2.68	2.89	3.06
	(0.41)	(0.46)	(0.47)	(0.49)	(0.49)	(0.50)
2 hours	1.56	2.43	2.95	3.31	3.58	3.81
	(0.56)	(0.63)	(0.66)	(0.68)	(0.69)	(0.70)
6 hours	1.93	2.65	3.06	3.32	3.53	3.70
	(0.92)	(1.02)	(1.07)	(1.10)	(1.13)	(1.14)
1 day	2.26	3.13	3.61	3.93	4.17	4.37
	(1.62)	(1.89)	(2.02)	(2.10)	(2.16)	(2.21)
USD/CHF						
30 minutes	1.31	2.09	2.56	2.88	3.13	3.34
	(0.36)	(0.41)	(0.43)	(0.45)	(0.46)	(0.47)
1 hour	1.45	2.23	2.69	2.99	3.23	3.43
	(0.49)	(0.53)	(0.56)	(0.57)	(0.58)	(0.59)
2 hours	1.70	2.61	3.13	3.49	3.76	3.99
	(0.66)	(0.74)	(0.77)	(0.79)	(0.81)	(0.82)
6 hours	2.14	2.96	3.42	3.73	3.96	4.16
	(1.08)	(1.20)	(1.26)	(1.30)	(1.32)	(1.34)
1 day	2.54	3.61	4.21	4.60	4.91	5.16
	(1.90)	(2.21)	(2.37)	(2.46)	(2.53)	(2.59)

Notes. The table gives extreme losses over different time intervals (30 minutes, 1 hour, 2 hours, 6 hours, 1 day) and corresponding to different loss probabilities (risk). The loss probabilities are expressed as once per a number of k years. The values within the parenthesis are the extreme losses computed for a Gaussian distribution with the mean and variance of the data. All values are given in percentages.

(respectively 0.0128 and 0.0161), together with the tail index estimate and sample variance $E[X^2] = 7 * 10^{-7}$ at the 30 minutes frequency, this implies $b = -0.108 * 10^{-3}$. Subsequently, one reuses the equality (15) to calculate the daily quantile as 0.0213. This is close to the number 0.0237 from Table 5 obtained by direct estimation (the difference may stem from the fact that the data are not independent).

6. Conclusion

More data does not always imply more information. The theme of this paper is that the high frequency forex data do provide substantial additional information concerning the risk on large losses. We showed that the power decline induces two types of benefits for the usage of high frequency data. The first benefit is that for the extreme losses, which are the focus of risk management, the risks are to a first order approximation additive over time. In particular we obtained the α-root rule for extrapolating single period risk of loss calculations to multi-period horizons, which is as easy as the square root rule for the case of the normal model. This enables one to bypass elaborate reestimation on lower frequencies. Because typically the second moment is bounded, i.e. $\alpha > 2$, the assumption of normality gives too conservative loss estimates at the low frequency levels (and is overly optimistic at the highest frequencies). The efficiency and bias properties of the estimators generally improve thanks to the usage of high frequency large data sets. As a rule of thumb one would therefore only estimate the relevant parameters on the highest frequency data that are available, and subsequently extrapolate the results to the lower frequencies without reestimation.

Appendix

We give a proof for the second order convolution Lemma 4. In addition to the assumptions in the main text, recall $\alpha > 2$, $\beta > 0$, we also assume that the monotone density theorem holds, such that the tails of the density of $F(x)$ satisfy

$$f(\pm x) = \alpha a x^{-\alpha-1} + ab(\alpha+\beta)x^{-\alpha-\beta-1} + o(x^{-\alpha-\beta-1}), \quad \text{as } x \to \infty.$$

Proof: Divide the area over which one has to integrate into five parts A, B, C, D and E; where $P\{A\} = P\{X_1 + X_2 \le s, \quad X_1 > -\frac{s}{2}, \quad X_2 > \frac{s}{2}\}$, $P\{C\} = P\{X_1 \le \frac{s}{2}, \quad X_2 \le \frac{s}{2}\}$, $P\{D\} = P\{X_1 + X_2 \le s, X_1 \le -\frac{s}{2}, X_2 > \frac{s}{2}\}$, and where $P\{B\}$ and $P\{E\}$ are the symmetric counterparts of $P\{A\}$ and $P\{D\}$ respectively. By the symmetry assumption $P\{D\} = P\{E\}$ and $P\{B\} = P\{A\}$ as $s \to \infty$.

We calculate $P\{C\}$, $P\{A\}$ and $P\{D\}$ and start by $P\{C\}$:

$$\begin{aligned} P\{C\} &= F^2\left(\frac{s}{2}\right) \\ &= 1 - 2a\left(\frac{s}{2}\right)^{-\alpha} - 2ab\left(\frac{s}{2}\right)^{-\alpha-\beta} + a^2\left(\frac{s}{2}\right)^{-2\alpha} + o(s^{-\alpha-\beta}). \end{aligned}$$

as $s \to \infty$.

The probability $P\{A\}$ takes more

$$\begin{aligned} P\{A\} &= \int_{-s/2}^{s/2} \left[F_2(s-x) - F_2\left(\frac{s}{2}\right)\right] f_1(x)\mathrm{d}x \\ &= \int_{-s/2}^{s/2} F_2(s-x) f_1(x)\mathrm{d}x - \int_{-s/2}^{s/2} F_2\left(\frac{s}{2}\right) f_1(x)\mathrm{d}x = I - II, \end{aligned}$$

say. For integral I note that Young's form of a second order Taylor expansion gives

$$(s-x)^{-\alpha} = s^{-\alpha} + \alpha s^{-\alpha-1}x + \frac{(\alpha+1)\alpha}{2}s^{-\alpha-2}x^2 + o(s^{-\alpha-2}x^2).$$

Hence, for large s and recalling the assumptions $\alpha > 2$, $\beta > 0$

$$\begin{aligned} I =& \left[1 - as^{-\alpha} - abs^{-\alpha-\beta}\right]\left[1 - 2a\left(\frac{s}{2}\right)^{-\alpha} - 2ab\left(\frac{s}{2}\right)^{-\alpha-\beta}\right] \\ &- \alpha as^{-\alpha-1}E[X] - \frac{(\alpha+1)\alpha}{2}as^{-\alpha-2}E\left[X^2\right] \\ &+ o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}), \end{aligned}$$

where terms like $(\alpha+\beta)abs^{-\alpha-\beta-1}E[X]$ and $\alpha^2(\alpha-1)^{-1}a^2 2^\alpha s^{-2\alpha}$ are dropped, since these are of smaller order. For part II

$$\begin{aligned} II =& F_2\left(\frac{s}{2}\right)\int_{-s/2}^{s/2} f_1(x)\mathrm{d}x \\ =& \left[1 - a\left(\frac{s}{2}\right)^{-\alpha} - ab\left(\frac{s}{2}\right)^{-\alpha-\beta} + o(s^{-\alpha-\beta})\right] \\ &\times\left[1 - 2a\left(\frac{s}{2}\right)^{-\alpha} - 2ab\left(\frac{s}{2}\right)^{-\alpha-\beta} + o(s^{-\alpha-\beta})\right]. \end{aligned}$$

Combine the two parts to obtain $P\{A\}$:

$$\begin{aligned} P\{A\} =& -as^{-\alpha} - abs^{-\alpha-\beta} + a\left(\frac{s}{2}\right)^{-\alpha} + ab\left(\frac{s}{2}\right)^{-\alpha-\beta} \\ &- \alpha as^{-\alpha-1}E[X] - \frac{(\alpha+1)\alpha}{2}as^{-\alpha-2}E\left[X^2\right] \\ &+ o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}). \end{aligned}$$

The last probability that we need is

$$P\{D\} = \int_{-\infty}^{-s/2}\left[F_2(s-x) - F_2\left(\frac{s}{2}\right)\right]f_1(x)\mathrm{d}x.$$

Since for $x \leq -s/2$ it is immediate that the integrand satisfies

$$F_2(s-x) - F_2\left(\frac{s}{2}\right) \leq a\left(\frac{s}{2}\right)^{-\alpha} + ab\left(\frac{s}{2}\right)^{-\alpha-\beta}.$$

Hence

$$P\{D\} \leq \left[a\left(\frac{s}{2}\right)^{-\alpha} + ab\left(\frac{s}{2}\right)^{-\alpha-\beta}\right] F_2\left(-\frac{s}{2}\right) = O(s^{-2\alpha}).$$

Adding up we find that

$$\begin{aligned} P\{X_1 + X_2 \ \leq \ s\} &\approx P\{C\} + 2P\{A\} \\ &= 1 - 2as^{-\alpha}\left(1 + bs^{-\beta} + \alpha E[X]s^{-1} + \frac{\alpha(\alpha+1)}{2}E\left[X^2\right]s^{-2}\right) \\ &\quad + o(s^{-\alpha-2}) + o(s^{-\alpha-\beta}). \end{aligned}$$

□

Acknowledgment

The work on this paper began when Wolfgang Härdle suggested to use the subsample bootstrap technique. We benefitted from the insightful suggestions by Tim Bollerslev, Eric Ghysels, Namwon Hyung, Kees Koedijk, Jorgen Olsen, Liang Peng, J. Robert Ward and the participants of the Tinbergen Institute conference on ‘extreme value theory with applications’ in Rotterdam, the ESEM94 in Maastricht, and the conference on ‘extremes, risk and safety’ in Gothenburg. We are grateful for the detailed comments by two referees and support of the editor. The empirical work in this paper was accomplished when Dacorogna and Pictet were at Olsen and Associates. They would like to thank the company for the excellent research facilities it provided.

Notes

1. See Dowd (1998) for a standard treatment of position limits and VaR under normality.
2. We are grateful to a referee who induced us to provide a formal argument for our conjecture.
3. But these are often so small that they can be ignored. For the high intra day exchange rate returns the interest gains are per definition zero, since these are only imputed on overnight positions.
4. In the above propositions the benchmark assumption of normality is not crucial. For example Stahl (1997) has argued that a Chebyshev bound based VaR yields a justification for the Basle multiplication factor of 3. It can be shown, however, that if the returns are heavy tailed the Chebyshev bound also yields $\lim_{t\to\infty} P\{Y \leq -t\}/P\{X \leq -t\} = 0$, as in Proposition 1; see De Haan et al. (1994).
5. We are extremely grateful to Namwon Hyung for pointing this out and thereby correcting the earlier statement of the result.

6. The choice of the parameter values reflects the typical values reported in the literature, see Baillie and McMahon (1989, ch. 4). Note that the AR coefficient is close to 1, where the ARCH process is no longer covariance stationary. But the process is still stationary even for larger values (when $\alpha < 2$), see e.g. Groenendijk et al. (1995).
7. We experimented with different values for ε, but this did not make much difference.
8. Note that the stable density and the Student density are continuous functions of respectively the characteristic exponent and the degrees of freedom, and hence converge to the normal density uniformly.
9. This assumes that the returns are i.i.d.

References

Baillie, R.T. and Dacorogna, M.M., ''Special issue on high frequency data,'' *Journal of Empirical Finance* 4, 69–293, (1997).

Baillie, R.T. and McMahon, P.C., *The Foreign Exchange Market*, Cambridge University Press, Cambridge, 1989.

Boothe, P. and Glassman, D., ''The statistical distribution of exchange rates, empirical evidence and economic implications,'' *Journal of International Economics* 22, 297–319, (1987).

Dacorogna, M.M., Müller, U.A., Nagler, R.J., Olsen, R.B., and Pictet, O.V., ''A geographical model for the daily and weekly seasonal volatility in the FX market,'' *Journal of International Money and Finance* 12(4), 413–438, (1993).

Danielsson, J., de Haan, L, Peng, L., and de Vries, C.G., ''Using a bootstrap method to choose the sample fraction in tail index estimation,'' *Journal of Multivariate Analysis* 76, 226–248, (2001).

Danielsson, J. and de Vries, C.G., ''Beyond the sample: Extreme quantile and probability estimation with applications to financial data,'' *Tinbergen Institute*, discussion paper, TI98-016/2, 1998.

De Haan, L., Jansen, D.W., Koedijk, K.G., and De Vries, C.G., ''Safety first portfolio selection, extreme value theory and long run asset risks'' In: J. Galambos, J. Lechner, and E. Simiu, (eds), *Extreme Value Theory and Applications*, Kluwer, Dordrecht, 471–488, (1994).

De Haan, L., Resnick, S.I., Rootzén, H., and De Vries, C.G., ''Extremal behavior of solutions to a stochastic difference equation with application to ARCH processes,'' *Stochastic Processes and their Applications* 32, 213–224, (1989).

Dowd, K., *Beyond Value at Risk*, Wiley, Chichester, 1998.

Drees, H. and Kaufman, E., ''Selection of optimal sample fraction in univariate extreme value estimation,'' *Stochastic Processes and their Applications* 75, 149–172, (1998).

Fama, E.F. and Miller, M.H., *The Theory of Finance*, Dryden Press, Hinsdale, 1972.

Feller, W., *An Introduction to Probability Theory and its Applications*, Volume II, John Wiley, New York, (second edition), 1971.

Geluk, J., de Haan, L., Resnick, S., and Starica, C., ''Second order regular variation, convolution, and the central limit theorem,'' *Stochastic Processes and their Applications* 69, 139–159, (1997).

Goldie, C.M. and Smith, R.L., ''Slow variation with remainder: Theory and applications,'' *Quarterly Journal of Mathematics*, Oxford 2nd series, 38, 45–71, (1987).

Goldie, C.M., ''Implicit renewal theory and tails of solutions of random equations,'' *The Annals of Applied Probability* 1, 126–166, (1991).

Groenendijk, P.A., Lucas, A., and de Vries, C.G., ''A note on the relationship between GARCH and symmetric stable processes,'' *Journal of Empirical Finance* 2, 253–264, (1995).

Hall, P., ''On some simple estimates of an exponent of regular variation,'' *Journal of the Royal Statistical Society* Series B, 44, 37–42, (1982).

Hall, P., ''Using the bootstrap to estimate mean square error and select smoothing parameter in nonparametric problem,'' *Journal of Multivariate Analysis* 32, 177–203, (1990).

Hill, B.M., ''A simple general approach to inference about the tail of a distribution,'' *The Annals of Statistics* 19, 1547–1569, (1975).

Hols, M.C. and De Vries, C.G., ''The limiting distributions of extremal exchange rate returns,'' *Journal of Applied Econometrics* 6, 287–302, (1991).

Leadbetter, M.R., Lindgren, G., and Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, Berlin, 1983.

Müller, U.A., Dacorogna, M.M., Olsen, R.B., Pictet, O.V., Schwarz, M., and Morgenegg, C., ''Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis,'' *Journal of Banking and Finance* 14, 1189–1208, (1990).

Stahl, G., ''Three cheers,'' *Risk* 10, 67–69, (1997).